J|A|C|S
COMMUNICATIONS

Gas-Phase Folding of a Two-Residue Model Peptide Chain: On the Importance of an Interplay between Experiment and Theory

E. Gloaguen,[†] B. de Courcy,[‡] J.-P. Piquemal,[*,‡] J. Pilmé,[‡] O. Parisel,[‡] R. Pollet,[§] H. S. Biswal,[†] F. Piuzzi,[†] B. Tardivel,[†] M. Broquier,[||] and M. Mons[*,†]

CEA, IRAMIS, Service des Photons, Atomes et Molécules (SPAM), and CNRS, Laboratoire Francis Perrin (LFP), URA 2453, 91191 Gif-sur-Yvette, France, CEA, IRAMIS, Service Interdisciplinaire sur les Systèmes Moléculaires et les Matériaux (SIS2M), 91191 Gif-sur-Yvette, France, CNRS, Institut des Sciences Moléculaires d'Orsay (ISMO), FRE 3363, and Université Paris-Sud, CLUPS, 91401 Orsay, France, and UPMC, Univ. Paris 06, Laboratoire de Chimie Théorique and CNRS, UMR 7616, 75252 Paris, Cedex 05, France

Received May 12, 2010; E-mail: jpp@lct.jussieu.fr; michel.mons@cea.fr

Abstract: In order to assess the ability of theory to describe properly the dispersive interactions that are ubiquitous in peptide and protein systems, an isolated short peptide chain has been studied using both gas-phase laser spectroscopy and quantum chemistry. The experimentally observed coexistence of an extended form and a folded form in the supersonic expansion was found to result from comparable Gibbs free energies for the two species under the high-temperature conditions ($\leq$320 K) resulting from the laser desorption technique used to vaporize the molecules. These data have been compared to results obtained using a series of quantum chemistry methods, including DFT, DFT-D, and post-Hartree−Fock methods, which give rise to a wide range of relative stabilities predicted for these two forms. The experimental observation was best reproduced by an empirically dispersion-corrected functional (B97-D) and a hybrid functional with a significant Hartree−Fock exchange term (M06-2X). In contrast, the popular post-Hartree−Fock method MP2, which is often used for benchmarking these systems, had to be discarded because of a very large basis-set superposition error. The applicability of the atomic counterpoise correction (ACP) is also discussed. This work also introduces the mandatory theoretical examination of experimental abundances. ΔH(0 K) predictions are clearly not sufficient for discussion of folding, as the conformation inversion temperature is crucial to the conformation determination and requires taking into account thermodynamical corrections (ΔG) in order to computationally isolate the most stable conformation.

Because of the exponential increase in computational power, molecular modeling techniques are now able to address various experimental problems from material sciences to biology. The problem of protein folding remains a difficult one, however, as the complex physics at play require an accurate treatment of noncovalent weak interactions, such as induction and dispersion.[1] In this context, quantum chemistry (QC) methods[2] play a central role, potentially offering access to chemical accuracy for small model peptides and therefore providing useful reference computations for the design of modern tools, such as polarizable force fields,[3] that target large systems like proteins. For such benchmarking purposes, two different strategies are currently applied. On the one hand, ab initio post-Hartree−Fock methods are often assumed to be the most suitable because of their high accuracy, but despite their sophistication and their ability to account for dispersion interactions, their limits have been discussed in the literature,[2,4,5] in particular when considering the interaction of electron-rich groups such as phenyl rings with other parts of the molecule. Alternatively, density functional theory (DFT) methods are computationally more efficient, making them more suitable to larger systems, and a wealth of functionals have been recently developed to address the above-mentioned issues and in particular to account for electronic correlation.[6,7] In parallel with these theoretical developments, current gas-phase methods now enable experimentalists to prepare isolated biomolecules and optically characterize the most stable conformers observed.[8−11] By comparison with a library of spectroscopic features or with QC-calculated IR spectra on a selected set of structures, the experimentally observed conformers can be reliably assigned, at least in terms of H-bonding networks and secondary structure. Capitalizing on the state-of-the-art studies in both fields, we have carried out a joint experimental/theoretical investigation of a flexible test molecule aimed at an assessment of emblematic QC techniques. The two-residue Ala-based model peptide chosen, Ac-(Ala)$_2$-O-Bzl (where Ac- and -Bzl stand for *N*-acetyl and benzyl, respectively), is short enough that it can still be tractable theoretically but still contains both a phenyl ring and a flexible chain that can lead to dispersive effects and conformational diversity.

The most stable structures of Ac-(Ala)$_2$-O-Bzl were detected experimentally under isolated conditions using UV and IR/UV spectroscopy. The molecules were vaporized by laser desorption (second harmonic of a Nd:YAG laser) from a pressed powder sample mixed with micrometric graphite powder used as a matrix.[12] Vaporization took place downstream of the nozzle aperture into a pulsed supersonic expansion of a He/Ne mixture or pure Ar (stagnation pressure 5−10 bar). The molecules were efficiently cooled in the jet, eventually leading to conformer-resolved UV features and resolved IR spectra. The near-UV spectrum of Ac-(Ala)$_2$-O-Bzl, as obtained by the mass-resolved resonant two-photon ionization (R2PI) technique (Figure 1a), exhibits two major sets of spectral features (origin band of the $\pi-\pi^*$ transition of the benzyl chromophore), namely, a first series of bands labeled A and a weaker and narrow isolated band labeled B. Double-resonance IR/UV spectroscopy showed that the NH-stretch IR spectra obtained from the A bands were identical and differed from those obtained from band B. This enabled us to assign the UV spectrum to two main conformers, labeled A and B. The IR spectrum of A (Figure 1b) is composed

---

[†] IRAMIS/SPAM and LFP.
[‡] UPMC, Univ. Paris 06.
[§] IRAMIS/SIS2M.
[||] ISMO and CLUPS.

10.1021/ja103996q

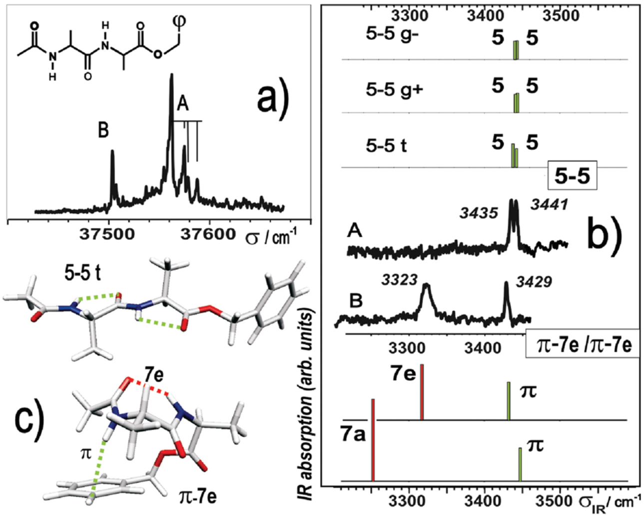

Figure 1. (a) Near-UV R2PI spectrum of the Ac-$(Ala)_2$-O-Bzl molecule, which exhibits features (labeled A and B) of two conformers. (b) Corresponding single-conformation IR spectra obtained by IR/UV double-resonance spectroscopy in the NH stretch region together with calculated scaled (0.9744)[13] harmonic frequencies of five conformations corresponding to the **5−5**, ***π*−7e**, and ***π*−7a** bonding schemes. (c) Bonding schemes and RI-B97-D/TZVPP structures that account for the IR spectra.

of a doublet in the 3440 cm^{-1} region. The B conformer presents one broad band at 3323 cm^{-1}, characteristic of a H–bond, together with an isolated band at 3429 cm^{-1}. As NH-stretch spectroscopy is a sensitive probe of the local environment of the NH bonds, this suggests that the H–bonding interactions of conformers A and B are qualitatively different: B alone contains a strong intramolecular H–bond. Comparison with previous reports on gas-phase peptide spectroscopy[10] suggests that the bands in the 3430−3440 cm^{-1} region, which are significantly shifted from the free NH region (3470 cm^{-1}), correspond to NH groups involved in weak interactions, namely, either C_5 or NH−π interactions.

In order to support this preliminary analysis, conformational explorations were performed [see the Supporting Information (SI)], followed by QC calculations at the DFT-D (RI-B97-D/TZVPP) level of theory, which has recently been shown to be an efficient assignment tool.[13] The most stable forms capable of accounting for the A spectrum correspond to the fully extended backbone conformations labeled **5−5** according to two consecutive C_5 interactions (Figure 1c). These forms (denoted as **5−5t**, **5−5g+**, and **5−5g−**) differ essentially by the orientation (trans, gauche+, and gauche−) of the benzyl tail relative to the backbone axis ($C_{CO}-O-C_{CH_2}-C_{Ar}$ dihedral). They were found to lie ca. 4 kcal/mol above the most stable ***π*−7e** folded conformation (see below). The good agreement between the experimental IR spectrum and the calculated results allows us to confirm the assignment of A as one of these extended-backbone **5−5** forms. The similarity of the three calculated spectra, however, forbids us from proposing a more precise assignment.

The lowest stable forms capable of accounting for the B spectrum correspond to compact folded structures stabilized by a C_7 interaction on the first Ala residue accompanied by a π interaction between the NH of the first Ala residue and the phenyl ring. Vibrational calculations enabled us to clearly distinguish the two types of C_7 bond (equatorial vs axial, according to the position of the Me side chain relative to the seven-membered ring; denoted as **7e** and **7a**, respectively). Figure 1c shows that the best agreement was obtained for the folded ***π*−7e** structure, which is the global minimum at the B97-D level of theory (the ***π*−7a** form was found to be less stable by 1.4 kcal/mol).

These results confirm our initial assignment. At this point, however, it is striking that both the A and B species were simultaneously observed, since as a result of its large stability relative to the **5−5** extended forms, the folded ***π*−7e** structure would be expected to be the only species observed. This paradoxical observation therefore demanded a critical examination of the experimental procedures:

First, possible distortion effects in the detection scheme due to conformer-dependent excited-state lifetimes[14] or ionization potentials (IP) had to be considered.[15] For this reason, two-color R2PI experiments with picosecond lasers, which enable a better ionization pump-up rate, were carried out (see the SI). Two-color experiments corresponding to excess energy up to 1.46 eV above the toluene IP did not provide any increase in the B feature. Since the oscillator strength of the $\pi-\pi^*$ transition in these conformers is essentially due to the benzyl moiety, it should therefore not greatly differ from one conformer to the other. One has to conclude from the experiments that both conformers were detected in the expansion at comparable abundances, with a slight advantage to the extended form.

Second, the issue of the origin of the observed conformational abundances had to be addressed. A and B strongly differ with respect to shape (extended and folded, respectively, with a significant barrier between them). One should therefore expect that the measured experimental abundances essentially reflect the conformational diversity prior to jet cooling. It is indeed admitted that conformational relaxation in the jet cannot enable the system to cross over barriers higher than a few kcal/mol and that the final conformational temperatures are usually better described by pre-expansion conditions.[16−20] In the case of laser desorption from a graphite surface (without jet cooling), relatively high vibrational temperature ranges (300−450 K) have been measured.[21] In the present case, the additional numerous collisions between the desorption plume and the early jet gas (pre-expansion $T \approx 320$ K) also probably contributed to some cooling, suggesting the use of the lower limit of the above-mentioned range (300 K) as a reference temperature, assuming a Boltzmann distribution. From this consideration, which is also in qualitative agreement with recent similar conclusions derived for a thermally vaporized flexible species,[20] the conformational populations in the jet should be considered as described by Gibbs free energies at 300 K rather than at 0 K. The comparable observed experimental abundances of A and B species suggested that 300 K Gibbs free energies for the extended and folded forms typically should not differ by more than 1 kcal/mol, therefore establishing a stringent assessment criterion for theoretical methods.

In order to carry out this assessment, the **5−5** and ***π*−7e** structures were optimized using a set of DFT functionals (in combination with the large 6-311++G** basis set), starting with PBE, a pure functional devoid of any exact Hartree−Fock (HF) exchange term, and continuing with hybrid functionals having an increasing empirical weight of HF exchange: PBE0, B3LYP, MPWB1K, and M06-2X.[22] With all of these methods, the three **5−5** forms were found to be quasi-isoenergetic (within 1 kcal/mol) at the electronic energy (ΔE_{el}) level (Table 1). However, their energies relative to the folded ***π*−7e** form ranged from ca. −2 kcal/mol for B3LYP to ca. +5 kcal/mol for M06-2X. It is important to notice that employing the previously mentioned B97-D/6-311++G** method, which includes a pure DFT functional embedding an explicit empirical dispersion term fitted on similar systems,[23] led to a close agreement with M06-2X.

Table 1. Electronic Energies (ΔE_{el}), ZPE-Corrected Energies [ΔH(0 K)], and Gibbs Free Energies [ΔG(300 K)] for the **5−5t**, **5−5g+**, and **5−5g−** Conformations Relative to ***π*−7e** Along with **5−5t**/***π*−7e** Inversion Temperatures (T_{inv}), As Obtained by Geometry Optimization at Different Levels of Theory Employing the 6-311++G** Basis Set[a]

	ΔE_{el}			ΔH(0 K)			ΔG(300 K)			
method	**5−5t**	**5−5g+**	**5−5g−**	**5−5t**	**5−5g+**	**5−5g−**	**5−5t**	**5−5g+**	**5−5g−**	T_{inv} (K)
PBE	0.14	−0.08	−0.23	−0.52	−0.51	−0.67	−3.43	−3.05	−3.37	none
PBE0	0.09	−0.15	−0.29	−0.58	−0.59	−0.73	−3.38	−2.94	−3.18	none
B3LYP	−1.75	−1.78	−1.86	−2.32	−2.12	−2.19	−4.28	−3.72	−3.79	none
MPWB1K	1.88	1.36	1.34	0.60	0.48	0.14	−2.47	−2.27	−3.40	90
M06-2X	5.57	4.76	4.49	4.74	4.05	4.01	2.21	1.75	3.05	510
B97-D	5.43	4.72	4.36	4.26	3.74	3.41	0.48	0.06	−0.08	330
MP2	9.02	8.14	7.57	7.95	7.15	6.62	4.13	3.68	3.36	∼800
sp MP2 CP	5.10	4.41	3.65	4.03	3.42	2.70	0.21	−0.05	−0.56	not calcd
sp MP2 ACP-4	2.61	2.26	2.02	1.54	1.27	1.07	−2.28	−2.20	−2.19	not calcd

[a] All energies are given in kcal/mol. ZPE-corrected energies and Gibbs free energies were obtained from harmonic frequencies. BSSE-corrected MP2 single-point (sp) calculations were performed using either the standard counterpoise (CP) technique or the intramolecular ACP-4 approach (see the text).

Addition of relative zero-point energies (ΔZPE) to ΔE_{el} did not significantly change this ordering, although one may notice a slight trend of the **5−5** structures toward stabilization relative to ***π*−7e**.

When the thermal effects (ΔG) are taken into account, however, one notices that whatever method is considered, the folded form is strongly destabilized with respect to the extended ones, in agreement with a significant entropic effect expected between such different conformational strains (see the SI for the temperature dependence). In the present case, the destabilization effect of the folded form is as large as 4 kcal/mol in magnitude, causing a new energetic ordering to emerge. Depending on the method used, very different **5−5** versus ***π*−7e** relative stabilities are now predicted. An alternative indicator of this stability order is the inversion temperature, T_{inv}, for which equal stabilities are predicted for **5−5t** and ***π*−7e** (Table 1). Examination of Table 1 suggests that the predicted values of both ΔG(300 K) and T_{inv} provide us with an acute assessment of the theoretical methods. Clearly, B97-D and to a lesser extent M06-2X provide the best agreement with experiment (exptl $T_{inv} \leq 320$ K). In contrast, PBE, PBE0, and MPWB1K, which behave similarly, lead to a strong underestimate of the ***π*−7e** stability; B3LYP performs still worse. One should notice that even when several ΔE_{el} DFT results seem to recover the experimental ordering, this is no longer true when all of the thermodynamic effects, especially the entropic term, are taken into account. Beyond B97-D, the results suggest that better agreement with experiment is found for hybrid functionals encompassing a high empirical weight of HF exchange.

At this stage of the discussion, it is now interesting to compare the DFT results with those obtained at the MP2 level of theory. It is worth noticing that the geometry change between the DFT- and MP2-optimized structures appears mainly between the benzene ring and the nearest methyl group for **5−5g+** and **5−5g−** and between the benzene ring and the two methyl groups adjacent to the NH−phenyl π H–bond in the ***π*−7e** structure (see Figure S3 in the SI); no significant geometry change was seen for **5−5t**. The MP2 minimization clearly enhances the π interaction. Therefore, relative to DFT, all of the conformers were found to be much less stable than ***π*−7e**, reaching a ΔE_{el} destabilization of ca. 8 kcal/mol. The inclusion of ZPE and thermodynamic effects slightly reduced the range of final energies (ΔG), but not enough to reproduce the experimental energetics, in agreement with the well-documented trend of the MP2 method to overestimate dispersion.[2,4,5] In this context, we then investigated the potential impact of the intramolecular basis-set superposition error (BSSE) using two types of calculations. First, with the counterpoise technique,[24] only the intermolecular BSSE between a benzene ring and an *N*-ethylacetamide group was considered, modeling the interacting moieties for each conformation. Table 1 shows that good agreement with the M06-2X and B97-D methods was obtained, suggesting that the failure of MP2 is a result of neglecting the BSSE correction. We then estimated the intramolecular BSSE correction using the newly introduced atomic counterpoise (ACP) approach, in which the BSSE is estimated as a sum of atomic contributions calculated as energy differences between a regular basis set and a subset consisting of basis functions on atoms separated by a minimum number of bonds.[25] We employed the ACP-*n* ($n = 4$) level,[25] where *n* stands for the number of bonds between atoms used in defining the basis-set subspace (see Appendix S5 in the SI). The results (Table 1) confirmed the trends observed with the intermolecular CP corrections of the MP2 level, whereas for DFT, small ACP corrections (found to be in the kcal/mol range) explained the good results obtained using B97-D and M06-2X.

To summarize, the interplay between conformation-resolved molecular jet experiments and quantum chemistry leads us to the following conclusions. First, it is essential to carefully consider three main effects when computing energies and conformations of such isolated peptides: (i) a good treatment of electronic correlation, which is required for handling dispersive interactions if they are not taken into account by empirical corrections; (ii) intramolecular BSSE corrections, which account for the use of limited basis sets; and (iii) thermodynamic corrections calculated at a temperature representative of the experimental conditions for a realistic comparison. Indeed, this work clearly introduces the mandatory theoretical examination of experimental abundances. ΔH(0 K) predictions are clearly not sufficient for discussion of folding, as the conformation inversion temperature is crucial to the conformation determination. Moreover, because of a large intramolecular BSSE, MP2 alone has been shown to be unsuitable for benchmarking energy surfaces and structures of dispersive systems such as Ac-(Ala)$_2$-O-Bzl, where energy differences between conformations range within 1 kcal/mol. Therefore, MP2 should not be taken as a reference method, especially in the design of new-generation force fields targeting a wide range of systems. Additional biologically relevant peptides are presently under investigation. In the future, BSSE-corrected MP2 could emerge as an ab initio reference "production" method, provided that remaining difficulties are fixed, namely, the system partition for ACP corrections, the limited applicability of MP2/ACP to large molecules because of its numerical cost, and the absence of analytical gradients for the inclusion of ACP/BSSE corrections in the geometry optimization procedure. For now, DFT functionals such as B97-D and to a lesser extent M06-2X perform relatively well thanks to limited BSSE effects. Beyond these popular methods, alternative strategies would be involve employing range-separated DFT functionals,[26] MP2 variants,[27] or extrapolations to the complete basis set limit. Such methods are currently being assessed in our groups and will be published in due course. Finally, it turns out that carefully chosen, well-balanced levels of theory seem to be reliable enough to provide an estimate of the conformational temperature from the measured experimental abundances.

Acknowledgment. Support from the French National Research Agency (ANR) is acknowledged (Grant ANR-08-BLAN-0158), as well as computer time allowances by CRIHAN (Rouen, France, Project 2008011) and GENCI (IDRIS, CINES, France, Project x2010075027). The authors thank Dr. R. Plowright for his careful reading of the manuscript.

Supporting Information Available: Picosecond two-color R2PI spectra (Appendix S1); DFT-D-calculated optimized structures, structural parameters, energetics, and vibrational frequencies of selected conformations (Appendix S2); *T*-dependent energetics (Appendix S3); comparison between MP2 and DFT-D geometries (Appendix S4); and ACP definition (Appendix S5). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Stone, A. J. *Theory of Intermolecular Forces*; Oxford University Press: Oxford, U.K., 2001.
(2) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723.
(3) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *J. Chem. Theory Comput.* **2007**, *3*, 1960.
(4) Sponer, J.; Sabat, M.; Burda, J. V.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1999**, *103*, 2528.
(5) van Mourik, T. *J. Chem. Theory Comput.* **2008**, *4*, 1610.
(6) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2001.
(7) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford Science Publications: Oxford, U.K., 1989.
(8) de Vries, M. S.; Hobza, P. *Annu. Rev. Phys. Chem.* **2007**, *58*, 585.
(9) Simons, J. P. *Mol. Phys.* **2009**, *107*, 2435.
(10) Chin, W.; Piuzzi, F.; Dimicoli, I.; Mons, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1033.
(11) Gerhards, M., Spectroscopy of Neutral Peptides in the Gas Phase. In *Principles of Mass Spectrometry Applied to Biomolecules*; Laskin, J., Lifshitz, C., Eds.; Wiley: Hoboken, NJ, 2006; pp 3−62.
(12) Piuzzi, F.; Dimicoli, I.; Mons, M.; Tardivel, B.; Zhao, Q. *Chem. Phys. Lett.* **2000**, *320*, 282.
(13) Gloaguen, E.; Pollet, R.; Piuzzi, F.; Tardivel, B.; Mons, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 11385.
(14) Mons, M.; Piuzzi, F.; Dimicoli, I.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. A* **2006**, *110*, 10921.
(15) Mons, M.; Dimicoli, I.; Piuzzi, F. *Int. Rev. Phys. Chem.* **2002**, *21*, 101.
(16) Ruoff, R. S.; Klots, T. D.; Emilsson, T.; Gutowsky, H. S. *J. Chem. Phys.* **1990**, *93*, 3142.
(17) Godfrey, P. D.; Brown, R. D. *J. Am. Chem. Soc.* **1998**, *120*, 10724.
(18) Reha, D.; Valdes, H.; Vondrasek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. S. *Chem.—Eur. J.* **2005**, *11*, 6803.
(19) Shubert, V. A.; Baquero, E. E.; Clarkson, J. R.; James, W. H.; Turk, J. A.; Hare, A. A.; Worrel, K.; Lipton, M. A.; Schofield, D. P.; Jordan, K. D.; Zwier, T. S. *J. Chem. Phys.* **2007**, *127*, 21.
(20) Baquero, E. E.; James, W. H.; Choi, T. H.; Jordan, K. D.; Zwier, T. S. *J. Phys. Chem. A* **2008**, *112*, 11115.
(21) Handschuh, M.; Nettesheim, S.; Zenobi, R. *J. Phys. Chem. B* **1999**, *103*, 1719.
(22) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
(23) Korth, M.; Grimme, S. *J. Chem. Theory Comput.* **2009**, *5*, 993.
(24) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
(25) Jensen, F. *J. Chem. Theory Comput.* **2010**, *6*, 100.
(26) Savin, A. On Degeneracy, Near-Degeneracy and Density Functional Theory. In *Recent Developments and Applications of Modern Density Functional Theory*; Seminario, J., Ed.; Theoretical and Computational Chemistry, Vol. 4; Elsevier: Amsterdam, 1996; pp 327−357.
(27) Reinhardt, P.; Piquemal, J.-P. *Int. J. Quantum Chem.* **2009**, *109*, 3259.

JA103996Q